EMPIRICAL CENTROID FICTITIOUS PLAY: AN APPROACH FOR DISTRIBUTED LEARNING IN MULTI-AGENT GAMES

BRIAN SWENSON[†*], SOUMMYA KAR[†] AND JOÃO XAVIER[⋆]

Abstract

The paper is concerned with distributed learning in large-scale games. The well-known fictitious play (FP) algorithm is addressed, which, despite theoretical convergence results, is impractical in large settings due to intense computation and communication requirements. An adaptation of the FP algorithm, designated as the empirical centroid fictitious play (ECFP), is presented. In ECFP players respond to the centroid of all players' actions rather than track and respond to the individual actions of every player. It is shown that ECFP is well suited for large-scale games and convergence of the algorithm in terms of average empirical frequency (a notion made precise in the paper) to a subset of the Nash equilibria, designated as the consensus equilibria, is proven under the assumption that the game is a potential game with permutation invariant potential function. Furthermore, a distributed formulation of the ECFP algorithm is presented in which players, endowed with a (possibly sparse) preassigned communication graph, engage in local, non-strategic information exchange to eventually agree on a common equilibrium. Convergence results are proven for the distributed ECFP algorithm. It is also shown that the methodology of distributed ECFP can be extended to prove convergence of traditional FP in a distributed setting.

Index Terms

Games, Distributed Learning, Fictitious Play, Nash Equilibria, Consensus

The work was partially supported by the FCT project [PEst-OE/EEI/LA0009/2011] through the Carnegie-Mellon/Portugal Program managed by ICTI from FCT and by FCT Grant CMU-PT/SIA/0026/2009.

[†]Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA (brian-swe@andrew.cmu.edu and soummyak@andrew.cmu.edu).

[⋆]Institute for Systems and Robotics (ISR), Instituto Superior Tecnico (IST), Technical University of Lisbon, Portugal (jxavier@isr.ist.utl.pt).

I. INTRODUCTION

We consider a scenario where n players are engaged in repeated play of a finite strategic form game, and player-to-player communication is restricted to a local neighborhood (defined according to a communication graph, possibly sparse) of each agent. The question of interest is, can agent behavior rules be assigned which ensure agents learn a Nash equilibrium (NE) strategy yet are practical in a large-scale setting? We focus on the well-known fictitious play (FP) algorithm. Originally introduced as a method to compute NE in two player games [1], FP has since been studied extensively to determine the class of games for which it can be shown to converge.[1] While the algorithm does not converge in all games ([2],[3],[4]), positive convergence results have been proven for certain classes of games ([5],[6],[7],[8]). Of particular relevance to this paper is a result demonstrating convergence of FP in games with an arbitrarily large number of players under the assumption of identical interests [7]. This theoretically promising result suggests that FP might be an ideal algorithm for some large-scale settings; however, the prohibitively demanding communication and computational requirements of the algorithm make any large-scale implementation highly impractical.

We suggest that a learning algorithm satisfy the following two criteria to be considered well-suited to large-scale games: (1) The algorithm is computationally tractable for individual agents; (2) The algorithm admits a distributed implementation.[2]

FP does not satisfy either criteria. Players executing the FP algorithm must track the past actions (empirical distribution) of all other players (violating 2) and respond to this empirical distribution by solving an optimization problem which is exponentially complex in terms of the number of players (violating 1).

The main objective of this paper (see also [9]) is to propose an adaptation of the FP algorithm that is well suited to learning NE in large-scale games. We divide our approach into two parts. First we propose an adaptation of FP where players respond to the centroid of the marginal empirical distributions (the centroid distribution), rather than track and respond to the entire tuple of independent marginal distributions. We call this algorithm empirical centroid fictitious play (ECFP). The advantages of this approach are a reduction in computational complexity resultant from the inherent symmetry, and a dramatic mitigation of the FP information tracking problem by requiring players to track only a single vector (the centroid

[1]In reference to FP or one of its variants, we use the term convergence to mean that the empirical frequency distribution of a FP process asymptotically converges to the set of Nash equilibria, a notion to be made precise in section II.

[2]By a distributed implementation we mean that inter-agent information exchange may be restricted to a (preassigned) local neighborhood of each agent.

distribution) which is invariant to the number of players in the game. ECFP is of independent interest but is not by itself suited to large-scale games since it violates the distributed criterion given above by requiring precise knowledge of the centroid distribution, information unattainable without global knowledge.

It is interesting to interpret this approach in light of the highly structured nature of large-scale games. To fully describe the utility function of a single agent in a game where all players have m actions, it is necessary to specify m^n payoff values, an enormous number of parameters. However, it has been observed that 'rare is the game that models an application of interest and yet lacks sufficient structure to be specified with a reasonable number of parameters' [10]. Example large-scale games that admit a compact representation include congestion games [11], symmetric games [12], anonymous games [13], and graphical games [14]. Given that large-scale games of interest tend to have a compact representation, the general question arises as to whether a learning algorithm applied to a highly structured large-scale game might not equilibrate using a compact representation of player behavior. ECFP can be seen as an initial foray into this domain, for which we have gained some positive results.

The second part of our approach is to assume players are endowed with an ancillary communication structure through which they can exchange any information they desire. We represent this structure using a graph $G = (V, E)$ where a vertex represents a player, and an edge represents the ability of a player to exchange information with a neighboring player. Note that in the majority of game theory literature, a graph structure denotes the ability of a player to *observe* the actions of a neighbor ([15],[16][17],[18]), not to exchange information, as in our approach. In distributed ECFP, players will use the ancillary communication network to exchange estimates of the centroid distribution. Information exchange is considered non-strategic; players do not try to manipulate the information they send for strategic gain.

While the assumption of this communication structure is made for algorithmic convenience, we note that it makes some degree of sense applied to social and economic settings as well. People engaged in repeated play (e.g., a daily commute) might certainly 'talk with their friends,' exchanging opinions on the average behavior of the aggregate in the hope of optimizing their next round of play.

Our main contributions are threefold:

Main Contribution 1: We present empirical centroid fictitious play (ECFP), a variant of FP which is well suited to large-scale games. We show that ECFP converges in terms of *average* empirical distribution to a subset of the mixed strategy Nash equilibria, which we call the consensus equilibria. Convergence results are proven for games with identical permutation-invariant utility functions and can be extended to the larger class of games known as potential games [19], with the restriction that the potential function be permutation invariant. We emphasize that the mode of convergence used in this

paper (convergence in *average empirical frequency*) is different from the more conventional convergence in empirical frequency.

The concept of a consensus equilibrium is closely related to that of a symmetric equilibrium. The existence of symmetric equilibrium in finite normal form games was first proven by Nash [20] in the same work where the concept of Nash equilibrium was originally presented. In general, a symmetric equilibrium is a Nash equilibrium that is invariant under automorphisms of the game. A consensus equilibrium, on the other hand, is a Nash equilibrium in which all players use an identical strategy. In the case of a symmetric game, the two concepts coincide.

Main Contribution 2: We present the distributed ECFP algorithm, an implementation of ECFP in which agent policy update depends only on local neighborhood information exchange. We prove convergence of the algorithm to the set of consensus equilibria. Moreover, this convergence guarantees that each agent obtains an accurate estimate of the limiting equilibrium strategy.

Main Contribution 3: We present the distributed FP algorithm, an implementation of the well-known FP algorithm where agent policy update depends only on local neighborhood information exchange. We prove convergence to the set of Nash equilibria for games with identical interests. Distributed FP is computationally equivalent to traditional FP and therefore inherits some of the drawbacks for large-scale implementation. However, we consider this contribution valuable because of the archetypal role of FP as a learning algorithm. This result shows that our methodology can be generalized and might allow for distributed implementation of other FP variants.

A. Related Work

An overview of the subject of learning in games is found in [21]. Many large-scale learning algorithms exist that are not based on FP, including no-regret algorithms ([22],[23]), aspiration learning [24], and other model-free approaches ([25],[26],[27]). These learning algorithms tend to be fundamentally different than FP in that they do not track past actions of other players.

Variants of FP have been proposed for two player games ([28],[29],[30]),[21]), generally aimed at improving various aspects of the two player algorithm (i.e., faster convergence, convergence in specific games, etc.).

Sampled FP [31] addresses the problem of computational complexity of FP in large-scale games by using a Monte Carlo method to estimate the best response. Although computationally simple in the initial steps of the algorithm, the number of samples required to ensure convergence grows without bound.

Dynamic FP [17] applies principles of dynamic feedback from control theory to improve the convergence properties of a continuous-time version of FP. The algorithm is shown to be stable around some

Nash equilibria where traditional FP is unstable. While the results generalize to multi-player games, there is no mitigation of the information gathering problem. In [32], a similar algorithm utilizing only payoff based dynamics is presented. Similar stability results are shown when the class of games is restricted to games with a pairwise utility structure.

Joint strategy FP [33] is shown to converge for generalized ordinal potential games. Players track the utility each of their actions would have generated in the previous round, and then use a simple recursion to update the predicted utility for each action in the subsequent round. Actions are chosen by maximizing the predicted utility. In joint strategy FP, the information tracking problem is mitigated by requiring agents to track only the information germane to the computation of the predicted utility for actions of interest. No information gathering scheme is explicitly defined; players are assumed to have full access to the necessary information at all times. In distributed ECFP, the information gathering scheme is explicitly defined via a preassigned (but arbitrary) communication graph and convergence results are demonstrated when interagent communication is restricted to local neighborhoods conformant to the graph.

There is a relationship between the ancillary communication structure presented in this paper and the class of state-based potential games [34],[35]. In such games, the action space is augmented with an additional state space. Payoffs are based on both the action and the state. The states may take on various interpretations; in [35] a player's state space consists of a value and a set of estimates of other players' values. Players are permitted to observe the state space of neighbors and update their estimates accordingly.

Along these same lines, in distributed ECFP we assume players have a personal estimate of the centroid distribution which they update by observing the estimates of neighbors. While there is an interesting relationship between the concepts, there are fundamental and important differences, the most important of which is the overarching objective of each work. In [34],[35] the primary objective is to design a distributed optimization framework amenable to existing game-theoretic learning algorithms, whereas the objective of this paper is to design a learning algorithm that is well suited to large-scale games.

The remainder of the paper is organized as follows: Section II sets up notation to be used in the subsequent development. The set of consensus equilibria is defined and the classical (centralized) FP algorithm is reviewed in the same section. Section III introduces ECFP as a low-information-overhead repeated-play alternative to FP for learning consensus equilibria in multi-agent games. A fully distributed implementation of the proposed ECFP, the distributed ECFP, in multi-agent scenarios in which agent information dynamics is restricted to communication over a preassigned sparse communication network is presented and analyzed in section IV. In section V we discuss generalizations of ECFP, including a

distributed implementation of the traditional FP algorithm. In section VI we demonstrate an application of distributed ECFP in a traffic routing scenario. Finally, section VII concludes the paper.

II. Preliminaries

Let Γ be a normal form game with a set of players $N = \{1, \ldots, n\}$. The set of actions, or pure strategies, for player i is given by $Y_i = \{1, 2, \ldots, m_i\}$, and the set of joint actions is given by $Y^n = \prod_{i=1}^n Y_i$. The utility function of player i is given by $u_i(y) : Y^n \to R$.

The set of mixed strategies for player i is given by $\Delta_i = \{p \in \mathbb{R}^{m_i} : \sum_{k=1}^{m_i} p(k) = 1\}$, the m_i-simplex. A mixed strategy $p_i \in \Delta_i$ may be thought of as a probability distribution from which player i samples to choose an action. In this context, a pure strategy may be thought of as a vertex of the probability simplex. With a slight abuse of notation, we denote the set of actions, or pure strategies, for player i, in this context, using the notation $A_i = \{e_1, e_1, \ldots e_{m_i}\}$, where m_i is the number of strategies available to player i, and e_j is the jth canonical vector in $\mathbb{R}^{m_i}$. The set of joint mixed strategies is given by $\Delta^n = \prod_{i=1}^n \Delta_i$, and the set of joint actions is given by $A^n = \prod_{i=1}^n A_i$. A joint mixed strategy is given by the n-tuple $(p_1, p_2, \ldots, p_n)$, $p_i \in \Delta_i$. In this paper, we often make the assumption that players use identical action spaces, in which case we drop the subscript on individual action spaces and write $\Delta = \Delta_i \; \forall i$, $A = A_i \; \forall i$, and $Y = Y_i = \{1, 2, \ldots, m\} \; \forall i$.

The mixed utility function for player i is given by the multilinear function $U_i(\cdot) : \Delta^n \to \mathbb{R}$.

$$U_i(p_1, \ldots, p_n) := \sum_{y \in Y} u_i(y) p_1(y_1) \ldots p_n(y_n). \tag{1}$$

Note that the mixed utility $U_i(p)$ may be interpreted as the expected value of $u_i(y)$ given that the players' mixed strategies are independent. For convenience the notation $U_i(p)$ will often be written as $U_i(p_i, p_{-i})$ where $p_i \in \Delta_i$ is the mixed strategy for player i, and p_{-i} indicates the joint mixed strategy for all players other than i. This paper will deal mostly with games with identical utility functions such that $U_i(p) = U_j(p) \; \forall i, j$; in such cases we drop the subscript and write $U(p) = U_i(p) \; \forall i$.

Let $\{a_i(t)\}_{t=1}^\infty$ be a sequence of actions for player i, where $a_i(t) \in A_i \; \forall t$. Let $\{a(t)\}_{t=1}^\infty$ be the associated sequence of actions $a(t) \in A^n$. Note that $a_i(t) \in \mathbb{R}^m$; when necessary, we denote the kth element of the vector $a(t)$ by $a(t, k)$. Let $q_i(t)$ be the normalized histogram (empirical distribution) of the actions of player i up to time t, i.e., $q_i(t) = \frac{1}{t} \sum_{s=1}^t a_i(s)$. Similarly, $q(t) = \frac{1}{t} \sum_{s=1}^t a(s)$ is the joint empirical distribution corresponding to the joint actions of the players up to time t. The sequence of distributions $\{q(t)\}_{t=1}^\infty$ is often called a belief sequence.

A mixed strategy p is a Nash equilibrium of Γ if, for each player i, $U_i(p) \geq U_i(g_i, p_{-i})\ \forall g_i \in \Delta_i$. We define the set of Nash equilibria as

$$K = \{p \in \Delta^n : U_i(p) \geq U_i(g_i, p_{-i})\ \forall g_i \in \Delta_i,\ \forall i\},$$

and the subset of consensus equilibria as

$$C = \{p \in K\ :\ p_1 = p_2 = \cdots = p_n\}.$$

The set of ε-Nash equilibria is given by

$$K_\varepsilon = \{p \in \Delta^n\ :\ U_i(p) + \varepsilon \geq U_i(g_i, p_{-i})\ \forall g_i \in \Delta_i,\ \forall i\}, \tag{2}$$

and the set of ε-consensus equilibria as

$$\begin{aligned} C_\varepsilon = \{p \in \Delta^n\ :\ U_i(p) + \varepsilon \geq U_i(g_i, p_{-i})\ \forall g_i \in \Delta_i,\ \forall i, \\ p_1 = p_2 = \cdots = p_n\}. \end{aligned} \tag{3}$$

The distance of a distribution $p \in \Delta^n$ from C is given by $d(p, C) = \inf\{\|p - g^\star\| : g^\star \in C\}$. Throughout the paper $\|\cdot\|$ denotes the standard $\mathcal{L}_2$ Euclidean norm unless otherwise specified. For $\delta > 0$ we denote the set

$$B_\delta(C) = \{p \in \Delta^n : p_1 = p_2 = \ldots = p_n, d(p, C) < \delta\}.$$

Unless stated otherwise, we will restrict attention to games with identical permutation-invariant utilities; formally, we assume:

A. 1. *All players use the same strategy space.*

A. 2. *The players' utility functions are identical and permutation invariant.*

Note that under these assumptions, the set of consensus equilibria is known to be nonempty [12].

Let

$$\bar{q}(t) = \frac{1}{n} \sum_{i=1}^{n} q_i(t)$$

be the average empirical distribution. Let $\bar{q}^n(t) = (\bar{q}(t), \bar{q}(t), \ldots, \bar{q}(t)) \in \Delta^n$ denote the mixed strategy where all players use the empirical average as their individual strategy. For convenience in notation we sometimes write $U(\bar{q}_i(t), \bar{q}_{-i})$, where the subscripts indicate the strategy $\bar{q}(t)$ is being used by player i or by all players except i respectively. The vector $\bar{q}(t)$ will be extremely important in the exposition of ECFP.

A. Fictitious Play

In fictitious play the best response of player i at time t is given by[3]

$$v_i(q(t)) := \max_{\alpha_i \in A_i} U(\alpha_i, q_{-i}(t)). \tag{4}$$

In other words this means that at time t each player chooses a best response by assuming that the empirical distributions of the other players accurately represent their respective mixed strategies. A fictitious play process is a sequence $\{a(t)\}_{t=1}^{\infty} \in A^n$ such that,

$$v_i(q(t)) = U(a_i(t+1), q_{-i}(t)), \quad \forall i.$$

A fictitious play process is said to converge in empirical frequency if $\lim_{t\to\infty} d(q(t), K) = 0$. In [7] it was shown that a fictitious play process converges in the above sense for games satisfying **A.1** – **A.2**.

III. Empirical Centroid Fictitious Play

The difficulty of implementing FP in large-scale distributed games can be understood by analyzing the FP best response calculation given in (4). The two major problems with FP are

(i) Each player must have access to the marginal empirical frequency distributions of the other $n-1$ players in order to compute a best response. It is impractical for each agent to track the actions of all other agents.

(ii) The computational complexity of computing the mixed utility given an n-dimensional probability density function grows exponentially with the size of the game.

The key idea of ECFP is a modification of the best response function which mitigates both of these problems. Consider a scenario where a player knows the structure of the game but does not have the ability to track the individual actions, $a_i(t)$, of any single player. Rather, a player is only able to track the *average* action, $\bar{a}(t) := \frac{1}{n}\sum_{i=1}^{n} a_i(t)$, of the collective and therefore has access only to the *average* empirical distribution, $\bar{q}(t) = \frac{1}{n}\sum_{i=1}^{n} q_i(t)$. In ECFP a player computes a best response by assuming the average empirical distribution accurately represents the mixed strategy of each of the other players and maximizes her utility accordingly. The best response in ECFP is given by

$$v_i^m(\bar{q}(t)) := \max_{\alpha_i \in A_i} U(\alpha_i, \bar{q}_{-i}(t)), \tag{5}$$

$v_i^m(p) : \Delta \to \mathbb{R}$. We use the superscript m to distinguish between the modified best response (5) and the traditional FP best response (4). In this scheme the information tracking problem of FP is mitigated

[3] A maximizing α_i in (4) exists since the A_i is assumed to be finite.

by requiring players to track only the centroid distribution, a vector whose dimension is invariant to the number of players in the game. The problem of computational complexity is resolved in a less direct manner. By exploiting the symmetry inherent in the ECFP best response calculation, the computation can often be greatly simplified. For example, in the distributed traffic routing scenario of section VI, the complexity of the best response calculation is reduced to constant time complexity in terms of the number of players.

In an ideal ECFP process, the best response calculation is given by (5). We consider a more general case where players do not have access to $\bar{q}(t)$ directly. Instead, player i has access to $\hat{q}_i(t) \in \mathbb{R}^m$, an approximation or estimate of $\bar{q}(t)$. Let $\varepsilon_i(t) = \|\hat{q}_i(t) - \bar{q}(t)\|$ be the error in player i's estimate. We make the following assumption about the decay rate of the error:

A. 3. $\varepsilon_i(t) = O(\frac{\log t}{t^r}),$ *for some* $r > 0$.

This particular decay rate will appear naturally in the analysis of ECFP in a distributed setting. A sequence of actions $\{a(t)\}_{t=1}^{\infty}$ is an *empirical centroid fictitious play process* if

$$v_i^m(\hat{q}_i(t)) = U\left(a_i(t+1), \hat{q}_{-i}(t)\right), \tag{6}$$

where $\hat{q}_{-i}(t)$ is the $(n-1)$-tuple $\hat{q}_{-i}(t) = (\hat{q}_i(t), \ldots, \hat{q}_i(t))$ and the initial action $a_i(1)$ is chosen arbitrarily for all i. We note that the traditional definition of the mixed utility $U(p)$, given in (1), is defined over the domain Δ^n. The restriction of the domain to Δ^n is not necessitated by the definition; rather, it is a byproduct of the traditional approach dealing only with mixed strategies $p \in \Delta^n$. The approximated empirical distribution $\hat{q}_i(t) \in \mathbb{R}^m$, however, is permitted to be outside the simplex, Δ, and may even take negative values. In this case we retain the definition of $U(p)$, given by (1), but extend the domain to the set of all n-tuples of vectors in $\mathbb{R}^m$. This adjustment of the traditional definition expands the domain to an unbounded set, but for practical purposes, we note that assumption **A.3** implies $\{\hat{q}_i(t)\}_{t\geq 0}$ belongs to a compact set.

In (6), each player best responds using $\hat{q}_i(t)$ (her personal estimate of $\bar{q}(t)$) as the assumed mixed strategy for the other $n-1$ players. In ECFP, players learn a strategy which is a consensus Nash equilibrium strategy. We summarize the result in the following Theorem:

Theorem 1. *Let* $\{a(t)\}_{t=1}^{\infty}$ *be an ECFP process such that* ***A.1** – **A.3** hold. Then* $d(\bar{q}^n(t), C) \to 0$ *as* $t \to \infty$.

Proof: Let $\bar{a}(t) = \frac{1}{n}\sum_{i=1}^{n} a_i(t)$, where $\bar{a}(t) \in \Delta, a_i(t) \in A$. Let $\bar{a}^n(t) \in \Delta^n$ be the n-tuple $(\bar{a}(t), \ldots, \bar{a}(t))$.

Note that for $t \geq 1$

$$\bar{q}^n(t+1) = \bar{q}^n(t) + \frac{1}{t+1}\left(\bar{a}^n(t+1) - \bar{q}^n(t)\right). \tag{7}$$

Using (7) we write

$$U(\bar{q}^n(t+1)) = U\left(\bar{q}^n(t) + \frac{1}{t+1}\left(\bar{a}^n(t+1) - \bar{q}^n(t)\right)\right).$$

Applying the multilinearity of $U(\cdot)$, we obtain

$$\begin{aligned} U(\bar{q}^n(t+1)) = U(\bar{q}^n(t)) &+ \frac{1}{t+1}\sum_{i=1}^{n} U\left(\bar{a}_i(t+1), \bar{q}_{-i}(t)\right) \\ &- \frac{1}{t+1}\sum_{i=1}^{n} U(\bar{q}_i(t), \bar{q}_{-i}(t)) + \zeta(t+1). \end{aligned}$$

where we have explicitly written the first order terms of the expansion and collected the remaining terms in $\zeta(t+1)$. Note that the number of second order terms in the above expansion is finite and the terms are uniformly bounded since $\max_{p\in\Delta^n}|U(p)| < \infty$. Hence, there exists a positive constant M (independent of t) large enough such that $|\zeta(t+1| \leq M.(t+1)^{-2}$ for all t. Thus,

$$\begin{aligned} U(\bar{q}^n(t+1)) \geq U(\bar{q}^n(t)) &+ \frac{1}{t+1}\sum_{i=1}^{n} U\left(\bar{a}_i(t+1), \bar{q}_{-i}(t)\right) \\ &- \frac{1}{t+1}\sum_{i=1}^{n} U(\bar{q}_i(t), \bar{q}_{-i}(t)) - \frac{M}{(t+1)^2}. \end{aligned}$$

The permutation invariance and multilinearity of $U(\cdot)$ permits a rearranging of terms. We use the notation $[a_j(t)]_i$ to indicate the action of player j at time t being played by player i.

$$\begin{aligned} &\sum_{i=1}^{n} U\left(\bar{a}_i(t+1), \bar{q}_{-i}(t)\right) \\ &= \sum_{i=1}^{n} U\left(\left[\frac{1}{n}\sum_{j=1}^{n} a_j(t+1)\right]_i, \bar{q}_{-i}(t)\right) \\ &= \sum_{i=1}^{n}\frac{1}{n}\sum_{j=1}^{n} U\left([a_j(t+1)]_i, \bar{q}_{-i}(t)\right) \\ &= \sum_{i=1}^{n}\frac{1}{n}\sum_{j=1}^{n} U\left([a_j(t+1)]_j, \bar{q}_{-j}(t)\right) \\ &= \sum_{j=1}^{n} U\left(a_j(t+1), \bar{q}_{-j}(t)\right). \end{aligned}$$

Thus,

$$\begin{aligned}
&U\left(\bar{q}^n(t+1)\right) - U\left(\bar{q}^n(t)\right) + \frac{M}{(t+1)^2} \\
&\geq \frac{1}{t+1}\sum_{i=1}^{n} U\left(a_i(t+1), \bar{q}_{-i}(t)\right) \\
&- \frac{1}{t+1}\sum_{i=1}^{n} U(\bar{q}_i(t), \bar{q}_{-i}(t)).
\end{aligned} \tag{8}$$

Let $L_i(t+1) = v_i^m\left(\hat{q}_i(t)\right) - U\left(a_i(t+1), \bar{q}_{-i}(t)\right)$. Substituting in $L_i(t+1)$, (8) becomes

$$\begin{aligned}
&U\left(\bar{q}^n(t+1)\right) - U\left(\bar{q}^n(t)\right) + \frac{M}{(t+1)^2} + \frac{1}{t+1}\sum_{i=1}^{n} L_i(t+1) \\
\geq&\frac{1}{t+1}\sum_{i=1}^{n}\left(v_i^m\left(\hat{q}_i(t)\right) - U\left(\bar{q}_i(t), \bar{q}_{-i}(t)\right)\right) \\
=&\frac{\alpha_{t+1}}{t+1},
\end{aligned} \tag{9}$$

where

$$\alpha_{t+1} := \sum_{i=1}^{n}\left(v_i^m\left(\hat{q}_i(t)\right) - U\left(\bar{q}_i(t), \bar{q}_{-i}(t)\right)\right).$$

Note that $U(\cdot)$ is multilinear and therefore locally Lipschitz continuous. As noted earlier, assumption **A.3** implies that $\{\hat{q}_i(t)\}_{t\geq 1}$ is contained in a compact subset of $\mathbb{R}^m$. Therefore, there exists a positive constant K (independent of t), such that $|U(a_i(t+1), \hat{q}_{-i}(t)) - U(a_i(t+1), \bar{q}_{-i}(t))| \leq K\|(a_i(t+1), \hat{q}_{-i}(t)) - (a_i(t+1), \bar{q}_{-i}(t))\|$, for all t. By assumption **A.3**, $\|\hat{q}_{-i}(t) - \bar{q}_{-i}(t)\| = O(\frac{\log t}{t^r})$, and hence $|U(a_i(t+1), \hat{q}_{-i}(t)) - U(a_i(t+1), \bar{q}_{-i}(t))| = O(\frac{\log t}{t^r})$, which, by (6), implies $L_i(t) = O(\frac{\log t}{t^r})$. In particular, $\sum_{t=2}^{T}\frac{L_i(t)}{t} < B$ is bounded above by some $B \in \mathbb{R}$ for all $T \geq 1$. Summing over $1 \leq t \leq T$ in (9),

$$\begin{aligned}
&U(\bar{q}^n(T+1)) - U(\bar{q}^n(1)) + \sum_{t=1}^{T}\frac{M}{(t+1)^2} + \sum_{t=1}^{T}\sum_{i=1}^{n}\frac{L_i(t+1)}{t+1} \\
&\geq \sum_{t=1}^{T}\frac{\alpha_{t+1}}{t+1}.
\end{aligned}$$

Note that $\sum_{t=1}^{T}\frac{M}{(t+1)^2}$ is summable; therefore all terms on the left hand side are bounded above for all $T \geq 1$, and hence it follows that

$$\sum_{t=2}^{T}\frac{\alpha_t}{t} < \overline{B}$$

is bounded above by some $\overline{B} \in \mathbb{R}$, for all $T \geq 2$. Let $\beta_{t+1} = \sum_{i=1}^{n}\left[v_i^m(\bar{q}(t)) - U\left(\bar{q}^n(t)\right)\right]$, and note that, by definition of $v_i^m(\cdot)$, $\beta_t \geq 0$ for all t. By Lemma 4 in the appendix, $|v_i^m(\hat{q}_i(t)) - v_i^m(\bar{q}(t))| = O\left(\frac{\log t}{t^r}\right)$.

Thus,

$$|\alpha_t - \beta_t| = O\left(\frac{\log t}{t^r}\right)$$

and hence by Lemma 5, $\sum_{t=2}^{T} \frac{\beta_t}{t} < \infty$ converges as $T \to \infty$. By Lemma 3 it then follows that

$$\lim_{T\to\infty} \frac{\beta_2 + \beta_3 + \ldots + \beta_T}{T} = 0.$$

Subsequently, by Lemma 6, we obtain for every $\varepsilon > 0$,

$$\lim_{T\to\infty} \frac{\#\{1 \le t \le T : \bar{q}^n(t) \notin C_\varepsilon\}}{T} = 0.$$

By Lemma 8, this is equivalent to

$$\lim_{T\to\infty} \frac{\#\{1 \le t \le T : \bar{q}^n(t) \notin B_\delta(C)\}}{T} = 0$$

for every $\delta > 0$. Finally, by Lemma 9, we obtain $d(\bar{q}^n(t), C) \to 0$ as $t \to \infty$. ■

We emphasize that Theorem 1 shows that the n-tuple of the *average* empirical distribution converges to C, that is, $d(\bar{q}^n(t), C) \to 0$. This is not the same as the more traditional definition of convergence in empirical frequency,

$$d(q(t), C) \to 0 \text{ as } t \to \infty. \tag{10}$$

The practical meaning of Theorem 1 is that players do in fact learn a consensus equilibrium strategy. It is true that each player i has access only to the distribution $\hat{q}_i(t)$. However, the tuple of these distributions $(\hat{q}_1(t), \hat{q}_2(t), \ldots, \hat{q}_n(t))$ also converges asymptotically to the set of consensus equilibria, i.e.,

$$d((\hat{q}_1(t), \hat{q}_2(t), \ldots, \hat{q}_n(t)), C) \to 0,$$

by **A.3**. Therefore, player i has direct access to her portion of the convergent joint strategy. Taken as a whole, player i learns a strategy which is a Nash (consensus) equilibrium with respect to the strategies learned by other players.

IV. Distributed ECFP

A. *Distributed Problem Formulation*

The result given in Theorem 1 is powerful in that it guarantees convergence to a Nash (consensus) equilibrium given only an estimate of the average empirical distribution. We wish to implement the algorithm in a distributed setting where information exchange is restricted to a local neighborhood of each agent.

Consider the following problem formulation: Players are engaged in an n-player repeated game. Players are endowed with an ancillary communication network which we represent using the graph

$G = (V, E)$, where a vertex represents a player, and an edge represents the ability of two players to exchange information. Players are permitted to exchange information with neighbors once per iteration of the repeated game. Note this implies there is only one time scale for both communication and game play. The information available to a player is restricted to knowledge of her own actions, and whatever information her neighbors choose to share with her. Information exchange is non-strategic; players do not try to manipulate the information they send for strategic gain.

We maintain assumptions **A.1** and **A.2** given in section II, and we add an additional assumption pertaining to the communication network.

A. 4. *The graph $G = (V, E)$ modeling the ancillary inter-agent communication network is connected.*[4]

The following two matrices are defined to facilitate a more compact description of the algorithm. Let

$$Q(t) := (q_1(t)\ q_2(t)\ \ldots\ q_n(t))^T \in \mathbb{R}^{n\times m}$$

be the matrix containing players' empirical distributions. Let $\hat{q}_i(t) \in \mathbb{R}^m$ be player i's estimate of $\bar{q}(t) \in \mathbb{R}^m$. Let

$$\hat{Q}(t) := (\hat{q}_1(t)\ \hat{q}_2(t)\ \ldots\ \hat{q}_n(t))^T \in \mathbb{R}^{n\times m}$$

be a matrix containing players' estimates of the average empirical distribution. Let $\hat{q}(t) \in \mathbb{R}^{n}$[5] be the n-tuple $(\hat{q}_1(t), \ldots, \hat{q}_n(t))$. The tuple $\hat{q}(t)$ will be important in distributed ECFP; in particular we will prove that $\hat{q}(t)$ converges to the set of consensus equilibria.

B. Distributed ECFP Algorithm

Initialize

(i) At time $t = 1$, each player i chooses an arbitrary initial action $a_i(1)$. The initial empirical distribution for player i is given by $q_i(1) = a_i(1)$. Player i initializes her local estimate of the empirical distribution as

$$\hat{q}_i(1) = \sum_{j\in\Omega_i\cup\{i\}} w_{ij} q_j(1) \tag{11}$$

where Ω_i is the set of neighbors of player i and w_{ij} is a weighting constant.

Iterate

(ii) At each time $t > 1$, player i computes the set of best responses using $\hat{q}_i(t)$ as the assumed mixed

[4] A graph is said to be connected if there exists a path (possibly multi-hop) between any pair of vertices.

[5] As noted in section III, the estimate $\hat{q}(t)$ may be outside the set Δ.

strategy for each of the $n-1$ other players. The next action

$$a_i(t+1) \in \{\arg\max_{\alpha_i \in A_i} U(\alpha_i, \hat{q}_{-i}(t))\} \tag{12}$$

is played according to the best response calculation. In the event of multiple pure strategy best responses, any of the maximizing actions in (12) may be chosen arbitrarily. The local empirical distribution $q_i(t+1)$ is updated to reflect the action taken, i.e.,

$$q_i(t+1) = q_i(t) + \frac{1}{t+1}(a_i(t+1) - q_i(t)).$$

(iii) Subsequently each player i computes a new estimate of the network-average empirical distribution using the following update rule:

$$\hat{q}_i(t+1) = \sum_{j \in \Omega_i \cup \{i\}} w_{i,j} \left(\hat{q}_j(t) + q_j(t+1) - q_j(t)\right), \tag{13}$$

where Ω_i is the set of neighbors of player i, and $w_{i,j}$ is a weighting constant.[6]
The update in (13) is represented in more compact notation as

$$\hat{Q}(t+1) = W\left(\hat{Q}(t) + Q(t+1) - Q(t)\right), \tag{14}$$

where $W \in \mathbb{R}^{n \times n}$ is a weighting matrix with entries $w_{i,j}$. We assume W satisfies the following assumption:

A. 5. *The weight matrix W is an $n \times n$ matrix that is doubly stochastic, aperiodic, and irreducible, with sparsity conforming to the communication graph G.*

Note that given assumption **A.4** (G is a connected graph), it is always possible to find a matrix W satisfying these conditions (see [36]).

C. Distributed ECFP: Main Result

We refer to any sequence of actions $\{a(t)\}_{t=1}^{\infty}$ which can be attained using the distributed ECFP algorithm of section IV-B as a *distributed ECFP process*. In a distributed ECFP process, players learn a consensus equilibrium strategy in a setting where information exchange is restricted to a local neighborhood of each agent. The result is summarized in the following Theorem.

[6]Note that the set $\Omega_i \cup \{i\}$ in the summation indicates that player i uses its own (local) information and that of her neighbors to update her estimate. The update rule is clearly distributed as information exchange is restricted to neighboring players only.

Theorem 2. *Let $\{a_i(t)\}_{t=1}^{\infty}$ be a distributed ECFP process such that assumptions **A.1**, **A.2**, **A.4**, and **A.5** hold. Then $d(\hat{q}(t), C) \to 0$ as $t \to \infty$. In particular, the agent estimates $\hat{q}_i(t)$'s reach asymptotic consensus, i.e. $d\left(\hat{q}_i(t), \hat{q}_j(t)\right) \to 0$ as $t \to \infty$ for each pair (i, j) of agents. Moreover, the agents achieve* asymptotic strategy learning, *in the sense that $d((\hat{q}_i(t))^n, C) \to 0$ as $t \to \infty$ for all $i = 1, \ldots, n$.*

This result implies that the n-tuple $(\hat{q}_1(t), \ldots, \hat{q}_n(t))$ converges to the set C; since $\hat{q}_i(t)$ is available to player i, player i learns the component of the consensus equilibrium strategy relevant to her.

Proof: We would like to apply the results of Theorem 1 to the distributed ECFP process. Assumptions **A.1** and **A.2** hold in a distributed ECFP process by assumption. By Lemma 2, the error in a distributed ECFP process decays as $\|\hat{q}_i(t) - \bar{q}(t)\| = O\left(\frac{\log t}{t}\right)$, thus **A.3** is satisfied (with $r = 1$), meeting all necessary assumptions for Theorem 1. Applying Theorem 1, $d(\bar{q}^n(t), C) \to 0$ as $t \to \infty$. By Lemma 2 we obtain, $\|\hat{q}_i(t) - \bar{q}(t)\| \to 0$ as $t \to 0$, and the result $d(\hat{q}(t), C) \to 0$ as $t \to \infty$ follows. ∎

Again, we emphasize that this mode of convergence is not the same as the more traditional convergence in empirical frequency, given in (10).

V. Generalizations

A. ECFP in Potential Games

The assumption **A.2** of identical permutation invariant utility functions can be relaxed in lieu of the following broader assumption:

A. 6. *The game Γ is an exact potential game with a permutation invariant potential function.*

A game Γ is an exact potential game if there exists some function $P(y) : Y^n \to \mathbb{R}$, such that

$$\begin{aligned} & u_i(y_i', y_{-i}) - u_i(y_i'', y_{-i}) \\ & = P(y_i', y_{-i}) - P(y_i'', y_{-i}) \; \forall i \in N, \forall y_i', y_i'' \in Y_i. \end{aligned}$$

The function $P(y)$ is called a potential function for Γ. The generalized form of Theorem 1 is as follows:

Theorem 3. *Let $\{a(t)\}_{t=1}^{\infty}$ be an ECFP process such that **A.1** (identical actions spaces), **A.3** ($\varepsilon_i(t) = O\left(\frac{\log t}{t^r}\right) \;\; \forall i$, for some $r > 0$), and **A.6** hold. Then $d(\bar{q}^n(t), C) \to 0$ as $t \to \infty$.*

Proof: Let $\Gamma_1 = (N, Y, \{U_i\}_{i \in N})$ be an exact potential game with potential function P. Let $\Gamma_2 = \left(N, Y, \{\tilde{U}_i\}_{i \in N}\right)$ be a game with the same set of players and actions as Γ_1, but with all players using P as their utility function ($\tilde{U}_i = P, \; \forall i$). Let C_{Γ_1} and C_{Γ_2} be the set of consensus equilibria in Γ_1 and Γ_2 respectively. Let $\bar{q}_{\Gamma_1}(t), \;\; \bar{q}_{\Gamma_2}(t)$ be the average empirical distributions corresponding to ECFP

processes in Γ_1 and Γ_2 respectively. Note that the set of consensus equilibria in Γ_1 and Γ_2 coincide [19]. Also note that Γ_1 and Γ_2 are best response equivalent [37], therefore a valid ECFP process for Γ_1 is a valid ECFP process for Γ_2, and vice versa. Γ_2 is a game with identical action spaces and identical permutation-invariant utility functions and therefore falls within the purview of Theorem 1. By Theorem 1, $d(\bar{q}^n_{\Gamma_2}(t), C_{\Gamma_2}) \to 0$. By best response equivalence, any valid ECFP process in Γ_1 is a valid ECFP process in Γ_2, therefore $d(\bar{q}^n_{\Gamma_1}(t), C_{\Gamma_2}) \to 0$. Since C_{Γ_1} and C_{Γ_2} coincide, $d(\bar{q}^n_{\Gamma_1}(t), C_1) \to 0$. ■

Potential games are studied in [19]. A game which admits an exact potential function is known as an exact potential game. The class of exact potential games includes congestion games [11]. Congestion games have many useful applications in economics and engineering. We present an example of a congestion game in the distributed traffic routing example presented in the applications section.

B. Distributed Implementation of Traditional FP

In the traditional FP best response (4), players are required to have precise knowledge of the empirical distribution of all other players at time t in order to pick the next stage action at time $t+1$. This requirement is tantamount to requiring that players have global knowledge and at first glance seems to disqualify the algorithm for distributed implementation. However, in this section we show that traditional FP can in fact be implemented in a distributed setting by using the same methodology employed to adapt ECFP for a distributed setting. The algorithm, which we call distributed FP (DFP), allows players to exchange information with neighbors via an ancillary communication network in order to estimate the empirical distribution of play. Players then pick a best response using the estimated empirical distribution and the traditional FP best response rule (4). The essential feature of the algorithm, the best response rule, is identical to that of FP; thus, the algorithm inherits the large-scale implementation problems of computational complexity and high information overhead associated with FP as mentioned in the introduction. Nevertheless, FP is an archetypal learning algorithm and demonstrating that our methodology extends to FP not only provides a rigorous demonstration of how this classic learning algorithm might be implemented in a distributed setting, but also suggests that a similar methodology might be employed to implement other FP variants in a distributed setting.

Consider the following problem setup: players are engaged in a repeated n-player game. Players are endowed with an ancillary communication network represented by the graph $G = (V, E)$, as described in section IV-A. Players are permitted to exchange information with neighbors once per iteration of the repeated game.

The key aspect of Theorem 1 which permits a distributed implementation of ECFP is embodied in assumption **A.3**. If the error in players' approximations of the average empirical distribution decays

quickly enough, then the process will converge to a consensus equilibrium. In the case of traditional FP we can show convergence given a similar assumption:

A. 7. *For all pairs* (i, j) *of players,* $\|\hat{q}^i_j(t) - q_j(t)\| = O\left(\frac{\log t}{t}\right)$*, where* $q_j(t)$ *is the empirical distribution of player* j*, and* $\hat{q}^i_j(t)$ *is the estimate which player* i *maintains of* $q_j(t)$*.*

Formally, we say the sequence $\{a(t)\}_{t=1}^{\infty}$ is an *asymptotically empirical FP process* if

$$\begin{aligned}&U(\hat{q}^i_1(t), \ldots \hat{q}^i_{i-1}(t), a_i(t+1), \hat{q}^i_{i+1}(t), \ldots, \hat{q}^i_n(t))\\&= v_i(\hat{q}^i_1(t), \ldots, \hat{q}^i_n(t))\end{aligned}$$

where $v_i(\cdot)$ is the FP best response defined in (4), and the initial action $a_i(1)$ is chosen arbitrarily for all i. We also make the following assumption, familiar from [7]:

A. 8. Γ *is a game with identical interests.*

Under these assumptions, a FP process can be shown to converge to the set of Nash equilibria, as stated in the following Theorem.

Theorem 4. *Let* $\{a(t)\}_{t=1}^{\infty}$ *be an approximately empirical FP process such that* ***A.7 - A.8*** *hold. Then* $d(q(t), K) \to 0$ *as* $t \to \infty$*.*

This can be seen as a generalization of the proof of the fictitious play property for games with identical interests found in [7]. The technical details of this proof follow very closely with the proof of Theorem 1, and are omitted here for brevity.

With this result in mind, we construct a distributed FP algorithm using the same methodology as the distributed ECFP algorithm of section IV-B. We show that the distributed FP algorithm converges in empirical frequency to the set of Nash equilibria.

1) Distributed FP algorithm: We introduce some notation to facilitate a compact description of the algorithm. Let $q_j(t) \in \mathbb{R}^{m_j}$ be the empirical distribution of player j. Let $\hat{q}^i_j(t) \in \mathbb{R}^{m_j}$ be the estimate which player i maintains for $q_j(t)$. Let $s = n\sum_{k \in N} m_k$ Let

$$\hat{q}^i(t) = ((\hat{q}^i_1(t))^T(\hat{q}^i_2(t))^T \ldots (\hat{q}^i_n(t))^T)^T \in \mathbb{R}^s$$

be the vector of player i's estimates. Let $q'_i(t)$ be an augmented vector representing the empirical distribution of player i such that

$$q'_i(t) = (0 \ldots (n \cdot q_i(t))^T \ldots 0)^T \in \mathbb{R}^s.$$

The augmented vector $q_i'(t)$ matches the general structure of $\hat{q}^i(t)$, but in the place of $q_i^i(t)$ we substitute in $n \cdot q_i(t)$ (a scaled copy of the true empirical distribution) and set all other entries to zero. Let

$$\hat{Q}(t) = \left(\hat{q}^1(t)\ \hat{q}^2(t)\ \cdots\ \hat{q}^n(t)\right)^T \in \mathbb{R}^{n\times s},$$

and let

$$Q'(t) = \left(q_1'(t)\ q_2'(t)\ \cdots\ q_n'(t)\right)^T \in \mathbb{R}^{n\times s}.$$

Initialize

(i) At time $t = 1$, each player i takes an arbitrary initial action $a_i(1)$. The initial empirical distribution for player i is given by $q_i(1) = a_i(1)$. Player i initializes her local estimate of the full empirical distribution as

$$\hat{q}^i(1) = \sum_{j\in\Omega_i\cup\{i\}} w_{ij} q'(1)$$

where Ω_i is the set of neighbors of player i and w_{ij} is a weighting constant.

Iterate

(ii) At each time $t > 1$, player i computes the set of best responses using $\hat{q}_j^i(t)$ as the assumed mixed strategy for player j. The next action

$$a_i(t+1) \in \\ \{\arg\max_{\alpha_i\in A_i} U\left(\hat{q}_1^i(t),\ldots,\hat{q}_{i-1}^i(t),\alpha_i,\hat{q}_{i+1}^i(t)\ldots,\hat{q}_n^i(t)\right)\} \tag{15}$$

is played according to the best response calculation. In the event of multiple pure strategy best responses, any of the maximizing actions in (15) may be chosen arbitrarily. The local empirical distribution $q_i(t+1)$ is updated to reflect the action taken, i.e.,

$$q_i(t+1) = q_i(t) + \frac{1}{t+1}(a_i(t+1) - q_i(t)).$$

(iii) Subsequently each player i computes a new estimate of the empirical distribution using the following update rule:

$$\hat{q}^i(t+1) = \sum_{j\in\Omega_i\cup\{i\}} w_{i,j}\left(\hat{q}^j(t) + q_j'(t+1) - q_j'(t)\right), \tag{16}$$

where Ω_i is the set of neighbors of player i, and $w_{i,j}$ is a weighting constant.[7]

[7]Note that the set $\Omega_i \cup \{i\}$ in the summation indicates that player i uses its own (local) information and that of her neighbors to update her estimate. The update rule is clearly distributed as information exchange is restricted to neighboring players only.

The update in (16) is represented in more compact notation as

$$\hat{Q}(t+1) = W\left(\hat{Q}(t) + Q'(t+1) - Q'(t)\right),$$

where $W \in \mathbb{R}^{n\times n}$ is a weighting matrix with entries $w_{i,j}$ and satisfying assumption **A.5**.

2) Convergence of Distributed FP: We refer to any sequence of actions $\{a(t)\}_{t=1}^{\infty}$ that can be attained using the distributed FP algorithm of section V-B1 as a *distributed FP process*. In distributed FP, players learn a Nash equilibrium strategy in a setting where information exchange is restricted to a local neighborhood of each player. The result is summarized in the following theorem.

Theorem 5. *Let $\{a(t)\}_{t=1}^{\infty}$ be a distributed FP process such that **A.4**, **A.5**, **A.7**, and **A.8** are satisfied. Then play converges in terms of empirical frequency to the set of Nash equilibria; that is, $d(q(t),K) \to 0$ as $t \to \infty$.*

Proof Sketch. *We would like to apply the results of Theorem 4 to the distributed FP process. Assumption **A.8** holds in a distributed FP process by assumption. Using a slight variation of Lemma 2, it can be shown that the error in a distributed FP process decays as $\|q_j^i(t) - q_j(t)\| = O\left(\frac{\log t}{t}\right)$ $\forall i,j \in N$; thus, **A.7** is satisfied, meeting all necessary assumptions for Theorem 4. Applying Theorem 4, $d(q(t),K) \to 0$ as $t \to \infty$.*

VI. Applications

A. Distributed Traffic Routing

We simulated distributed ECFP in a traffic routing scenario with 200 cars traversing a traffic network of five roads. This scenario is an instance of a congestion game. Any congestion game can be shown to satisfy assumption **A.6**, and therefore falls within the purview of ECFP. All vehicles start from the same location and have the same destination. The number of cars on road r for a joint strategy $y \in Y^n$ is given by $\sigma_r(y)$. The delay on road r for a joint strategy y is given by the cubic cost function

$$c_r(y) = a_3\sigma_r(y)^3 + a_2\sigma_r(y)^2 + a_1\sigma_r(y) + a_0,$$

where the coefficients $a_k \in \mathbb{R}$ are arbitrary. The utility for player i is given by $u_i(y) = -c_{y_i}(y)$, the negative of the travel delay experienced by player i. In this particular simulation, the weights $w_{i,j}$ (see (14)) were chosen according to the Metropolis-Hastings rule [38]. A graph of the communication network used for simulations is shown in Fig. 1. On the surface, the ECFP best response calculation (5) appears to have the same complexity as the FP best response calculation (4). However, the symmetry inherent

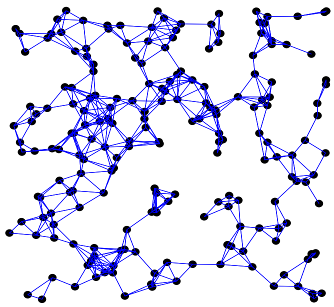

Fig. 1. Randomly generated sparse communication graph.

in the ECFP best response calculation can lead to simplifications. Given the cost functions used in this simulation, the ECFP best response can be simplified to an expression that is independent of the number of players. This fact is not unique to the cubic cost functions used here; it holds for games of this form with polynomial cost functions of any degree. The key results of ECFP are based around showing $d(\hat{q}(t), C) \to 0$. An important practical implication of this result is that $\hat{q}(t) \in K_{\varepsilon_t}$ where $\varepsilon_t \to 0$ and K_{ε_t} is an ε_t Nash equilibrium as defined in (2). Fig. 2 shows a plot of the minimum ε_t such that $q(t) \in K_{\varepsilon_t}$. This plot shows that ε_t tends to zero which is consistent with $d(\hat{q}(t), C) \to 0$, the claim of our main result in Theorem 2. For the traffic routing application, this means that players concurrently learn a mixed strategy ε_t consensus equilibrium (see (3)), where ε_t can be made arbitrarily small. Once ε_t is sufficiently small for the game designers' purposes, the algorithm can be terminated. The cost functions used to model road delay were specifically chosen as cubic polynomials in order to model a situation in which there may exist multiple consensus equilibria; distributed ECFP is particulary relevant to such situations since it can be used not only to compute a consensus equilibrium, but also to ensure that players agree on which consensus equilibrium is reached.

VII. Conclusions

We have introduced a variant of the well-known FP algorithm which we call empirical centroid fictitious play (ECFP). Rather than track and respond to the empirical distribution of each player, as in FP, ECFP tracks the centroid of the marginal empirical distributions and computes a best response with respect to this same quantity. The computational complexity of computing a best response is mitigated by the introduction of symmetry into the best response calculation, and the information tracking problem is mitigated by requiring players to track a quantity which is invariant to the number of players in the game. ECFP is shown to converge to the set of consensus equilibria, a subset of the Nash equilibria where all

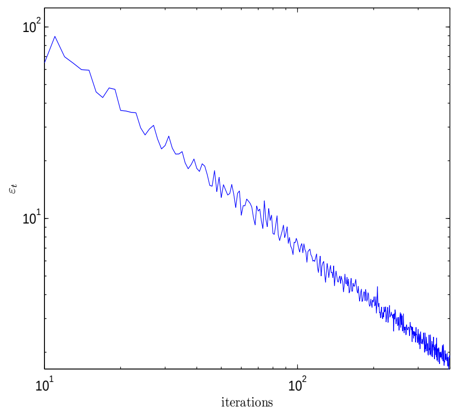

Fig. 2. The minimum ε_t on each iteration t such that $\hat{f}(t) \in K_{\varepsilon_t}$. The trend $\varepsilon_t \to 0$ is consistent with convergence to the set of consensus equilibria (i.e. $d(\hat{f}(t), C) \to 0$ as $t \to \infty$) as stated in Theorem 2.

players use an identical strategy, for potential games with permutation invariant potential functions.

We have provided a distributed implementation of ECFP which depends only on local information exchange, and we have proven convergence of the algorithm to the set of consensus equilibria. Furthermore, we have shown that the same approach can be used to formulate a distributed implementation of the traditional fictitious play algorithm.

An interesting future research direction would be to investigate if ECFP can be shown to converge to a consensus equilibrium for the more general class of symmetric games. It would also be of interest to investigate if ECFP can be shown to converge to the more general equilibrium concept of a symmetric equilibrium in games in which a consensus equilibrium may not exist.

REFERENCES

[1] G. W. Brown, *"Iterative Solutions of Games by Fictitious Play" In Activity Analysis of Production and Allocation*, T. Coopmans, Ed. New York: Wiley, 1951.

[2] L. S. Shapley, "Some topics in two-person games," *Advances in game theory*, vol. 52, pp. 1–29, 1964.

[3] J. S. Jordan, "Three problems in learning mixed-strategy Nash equilibria," *Games and Economic Behavior*, vol. 5, no. 3, pp. 368–386, Jul. 1993.

[4] D. P. Foster and H. P. Young, "On the nonconvergence of fictitious play in coordination games," *Games and Economic Behavior*, vol. 25, no. 1, pp. 79–96, Oct. 1998.

[5] J. Robinson, "An iterative method of solving a game," *The Annals of Mathematics*, vol. 54, no. 2, pp. 296–301, Sep. 1951.

[6] K. Miyasawa, "On the convergence of the learning process in a 2 x 2 non-zero-sum two-person game," DTIC Document, Tech. Rep., 1961.

[7] D. Monderer and L. S. Shapley, "Fictitious play property for games with identical interests," *Journal of Economic Theory*, vol. 68, no. 1, pp. 258–265, Jan. 1996.

[8] M. Benaım and M. W. Hirsch, "Mixed equilibria and dynamical systems arising from fictitious play in perturbed games," *Games and Economic Behavior*, vol. 29, no. 1, pp. 36–72, Oct. 1999.

[9] B. Swenson, S. Kar, and J. Xavier, "Distributed learning in large-scale multi-agent games: A modified fictitious play approach," in *46th Asilomar Conference on Signals, Systems, and Computers*, Pacifc Grove, CA, USA, Nov. 4 - 7 2012, pp. 1490 – 1495.

[10] C. H. Papadimitriou and T. Roughgarden, "Computing equilibria in multi-player games," in *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*. SIAM, 2005, pp. 82–91.

[11] R. W. Rosenthal, "A class of games possessing pure-strategy nash equilibria," *International Journal of Game Theory*, vol. 2, no. 1, pp. 65–67, 1973.

[12] S.-F. Cheng, D. M. Reeves, Y. Vorobeychik, and M. P. Wellman, "Notes on equilibria in symmetric games," in *Proceedings of Workshop on Game Theory and Decision Theory*, 2004.

[13] C. Daskalakis and C. Papadimitriou, "Computing equilibria in anonymous games," in *48th Annual IEEE Symposium on Foundations of Computer Science*, Oct. 2007, pp. 83–93.

[14] M. Kearns, M. L. Littman, and S. Singh, "Graphical models for game theory," in *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 2001, pp. 253–260.

[15] D. Gale and S. Kariv, "Bayesian learning in social networks," *Games and Economic Behavior*, vol. 45, no. 2, pp. 329–346, Nov. 2003.

[16] C. Eksin, P. Molavi, A. Ribeiro, and A. Jadbabaie, "Learning in linear games over networks," in *Proceedings of the 50th Annual Allerton Conference on Communications, Control, and Computing*, 2012.

[17] J. S. Shamma and G. Arslan, "Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria," *IEEE Transactions on Automatic Control*, vol. 50, no. 3, pp. 312–327, Mar. 2005.

[18] D. Rosenberg, E. Solan, and N. Vieille, "Informational externalities and emergence of consensus," *Games and Economic Behavior*, vol. 66, no. 2, pp. 979–994, Jul. 2009.

[19] D. Monderer and L. S. Shapley, "Potential Games," *Games and Economic Behavior*, vol. 14, no. 1, pp. 124–143, May 1996.

[20] J. Nash, "Equilibrium Points in n-person Games," *Proceedings of The National Academy of Sciences*, vol. 36, no. 1, pp. 48–49, 1950.

[21] D. Fudenberg and D. K. Levine, *The theory of learning in games*. MIT press, 1998, vol. 2.

[22] A. Jafari, A. Greenwald, D. Gondek, and G. Ercal, "On no-regret learning, fictitious play, and Nash equilibrium," in *Proceedings of the 18th International Conference on Machine Learning*, 2001, pp. 226–233.

[23] J. R. Marden, G. Arslan, and J. S. Shamma, "Regret based dynamics: convergence in weakly acyclic games," in *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*. ACM, 2007.

[24] G. C. Chasparis, A. Arapostathis, and J. S. Shamma, "Aspiration learning in coordination games," *SIAM Journal on Control and Optimization*, vol. 51, no. 1, pp. 465–490, Jan. 2013.

[25] J. R. Marden, H. P. Young, and L. Y. Pao, "Achieving pareto optimality through distributed learning," in *IEEE 51st Annual Conference on Decision and Control*. IEEE, 2012, pp. 7419–7424.

[26] B. S. Pradelski and H. P. Young, "Learning efficient Nash equilibria in distributed systems," *Games and Economic Behavior*, vol. 75, no. 2, pp. 882–879, Jul. 2012.

[27] J. R. Marden and J. S. Shamma, "Revisiting log-linear learning: Asynchrony, completeness and payoff-based implementation," *Games and Economic Behavior*, vol. 75, no. 2, pp. 788–808, Jul. 2012.

[28] D. Fudenberg and D. K. Levine, “Consistency and cautious fictitious play,” *Journal of Economic Dynamics and Control*, vol. 19, no. 5, pp. 1065–1089, Jul. 1995.

[29] D. S. Leslie and E. Collins, “Generalised weakened fictitious play,” *Games and Economic Behavior*, vol. 56, no. 2, pp. 285–298, Aug. 2006.

[30] A. Washburn, “A new kind of fictitious play,” *Naval Research Logistics*, vol. 48, no. 4, pp. 270–280, Jun. 2001.

[31] T. J. Lambert, M. A. Epelman, and R. L. Smith, “A fictitious play approach to large-scale optimization,” *Operations Research*, vol. 53, no. 3, pp. 477–489, May 2005.

[32] G. Arslan and J. S. Shamma, “Distributed convergence to Nash equilibria with local utility measurements,” in *Proceedings of the 43rd IEEE Conference on Decision and Control*, vol. 2, 2004, pp. 1538 – 1543.

[33] J. R. Marden, G. Arslan, and J. S. Shamma, “Joint strategy fictitious play with inertia for potential games,” *IEEE Transactions on Automatic Control*, vol. 54, no. 2, pp. 208–220, February 2009.

[34] J. R. Marden, “State based potential games,” *Automatica*, vol. 48, no. 12, pp. 3075–3088, Dec. 2012.

[35] N. Li and J. R. Marden, “Designing games for distributed optimization,” in *Proceedings of the 50th IEEE Conference on Decision and Control*, 2011.

[36] A. G. Dimakis, S. Kar, J. M. Moura, M. G. Rabbat, and A. Scaglione, “Gossip algorithms for distributed signal processing,” *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1847–1864, Nov. 2010.

[37] M. Voorneveld, “Best-response potential games,” *Economics letters*, vol. 66, no. 3, pp. 289–295, Mar. 2000.

[38] S. Chib and E. Greenberg, “Understanding the Metropolis-Hastings algorithm,” *The American Statistician*, vol. 49, no. 4, pp. 327–335, 1995.

[39] S. Rajagopalan and D. Shah, “Distributed averaging in dynamic networks,” *IEEE jornal of selected topics in signal processing*, vol. 5, no. 4, pp. 845–854, Aug. 2011.

[40] S. Kar and J. M. Moura, “Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs,” *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 4, pp. 674–690, Aug. 2011.

[41] J. Chen and A. H. Sayed, “Diffusion adaptation strategies for distributed optimization and learning over networks,” *Signal Processing, IEEE Transactions on*, vol. 60, no. 8, pp. 4289–4305, Aug. 2012.

[42] R. Olfati-Saber, J. A. Fax, and R. M. Murray, “Consensus and cooperation in networked multi-agent systems,” *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, Jan. 2007.

[43] D. Shah, “Gossip algorithms,” *Foundations and Trends in Networking*, vol. 3, no. 1, pp. 1–125, 2008.

[44] A. N. Shiryaev, *Probability*, ser. Graduate texts in mathematics. Springer-Verlag, New York, 1996, vol. 95.

VIII. APPENDIX

A. Distributed averaging in dynamic networks

This appendix concerns topics in distributed consensus in networks where node values are dynamic quantities. The results of this section are used to prove convergence of the distributed algorithms presented in sections IV-B and V-B1. Results in this section are similar to results on distributed averaging in networks with additive changes in node values and information dynamics in [39], [40], [41]. For a survey of traditional consensus and gossip algorithms, the reader may refer to [42],[43],[36].

Consider a network of n nodes connected through a communication graph $G = (V, E)$. The graph is assumed to be connected. Let $x_i(t) \in \mathbb{R}$ be the value of node i at time t, and let $x(t) \in \mathbb{R}^n$ be the vector

of values at all nodes. The goal is for each node to track the instantaneous average $\bar{x}(t) = \frac{1}{n}\sum_{i=1}^{n} x_i(t)$, $\bar{x}(t) \in \mathbb{R}$, given that the value at each node $x_i(t)$ is time varying. Let $\delta_i(t) = x_i(t+1) - x_i(t)$ be the change in the value at node i, and $\delta(t) = x(t+1) - x(t)$ be the vector of changes at all nodes, $\delta(t) \in \mathbb{R}^n$. Suppose the magnitude of the change at time t is bounded by $|\delta_i(t)| = |x_i(t+1) - x_i(t)| \leq \epsilon(t)\ \forall i$. We make the following assumption:

A. 9. *The sequence $\{\varepsilon(t)\}_{t=0}^{\infty}$ is monotone non-increasing.*

Let $\hat{x}_i(t) \in \mathbb{R}$ be the estimate of $\bar{x}(t)$ at node i and let $\hat{x}(t) \in \mathbb{R}^n$ be the vector of estimates. We make the following assumption pertaining to the initial error in players' estimates.

A. 10. $\hat{x}_i(0) - \bar{x}(0) = 0\ \forall i$.

Let the average be estimated using the update rule

$$\hat{x}(t+1) = W\left(\hat{x}(t) + x(t+1) - x(t)\right), \tag{17}$$

where the matrix $W \in \mathbb{R}^{n\times n}$ is aperiodic, irreducible and doubly stochastic with sparsity conforming to G. The following Lemma gives a bound for the error in the estimates of $\bar{x}(t)$.

Lemma 1. *Let the sequence $\{\hat{x}(t)\}_{t=1}^{\infty}$ be computed according to* (17) *such that assumptions **A.4**, **A.5**, and **A.10** hold and let the incremental change in $x(t)$ be bounded according to assumption **A.9**. Then the error at any time t is bounded by,*

$$\|\hat{x}(t) - \bar{x}(t)\mathbf{1}\| \leq \frac{2\sqrt{n}}{1-\lambda}\epsilon_{avg}(t),$$

where $\lambda = \sup_{y\in\mathbb{R}^n : \sum_i y_i = 0} \frac{\|Wy\|}{\|y\|}$, *and* $\epsilon_{avg}(t) = \frac{1}{t}\sum_{\tau=0}^{t-1}\epsilon(\tau)$ *is the time average of* $\{\epsilon(\tau)\}_{\tau=0}^{t-1}$.

Proof: Let $e(t) = \hat{x}(t) - \bar{x}(t)\mathbf{1}$ be the vector of errors in each players estimate of $\bar{x}(t)$, where $\mathbf{1}$ denotes the $n \times 1$ vector of all ones. Let

$$\bar{\delta}(t) = \frac{1}{n}\sum_i \delta_i(t),\ \ \forall t.$$

Using the relation (17) and the properties of doubly stochastic matrices, the vector of errors may be written recursively as,

$$e(t+1) = W\left(e(t) + \xi(t)\right) \tag{18}$$

where $\xi(t) = \delta(t) - \bar{\delta}(t)\mathbf{1}$. Note that

$$\begin{aligned}|\xi_i(t)| &= |\delta_i(t) - \bar{\delta}_i(t)| \\ &\leq |\delta_i(t)| + |\bar{\delta}_i(t)| \\ &\leq 2\epsilon(t),\end{aligned}$$

and

$$\|\xi(t)\|^2 = \sum_{i=1}^{n} (\xi_i(t))^2 \leq \sum_{i=1}^{n} 4\epsilon(t)^2 = 4n\epsilon(t)^2. \tag{19}$$

Using (18), the error $e(t)$ can be rewritten as a function of $\xi(t)$ and $e(0)$,

$$e(t+1) = \sum_{r=0}^{t} W^{r+1}\xi(t-r) + W^{t+1}e(0).$$

Using this relationship we establish an upper bound on the error,

$$\begin{aligned}\|e(t+1)\| &= \|\sum_{r=0}^{t} W^{r+1}\xi(t-r) + W^{t+1}e(0)\| \\ &\leq \sum_{r=0}^{t} \|W^{r+1}\xi(t-r)\| \\ &\leq \sum_{r=0}^{t} \lambda^{r+1}\|\xi(t-r)\|,\end{aligned} \tag{20}$$

where we have employed assumption **A.10**, $e(0) = 0$. Applying (19) in (20), we get

$$\|e(t+1)\| \leq \sum_{r=0}^{t} \lambda^{r+1} 2\sqrt{n}\epsilon(t-r).$$

Recall that $\varepsilon_{avg}(t) = \frac{1}{t}\sum_{\tau=0}^{t-1} \epsilon(\tau)$ is the time average of the sequence $\{\varepsilon(t)\}$ up to time t, and note that given our assumptions on W, it holds that $\lambda < 1$ (see [36]). Hence, by invoking Lemma 10 we have that

$$\begin{aligned}\|e(t+1)\| &\leq \sum_{r=0}^{t} \lambda^{r+1} 2\sqrt{n}\varepsilon_{avg}(t+1) \\ &= \left(\lambda \frac{1-\lambda^{t+1}}{1-\lambda}\right) 2\sqrt{n}\epsilon_{avg}(t+1) \\ &\leq \frac{2\sqrt{n}}{1-\lambda}\epsilon_{avg}(t+1).\end{aligned}$$

This gives us the desired upper bound for the error,

$$\|\hat{x}(t) - \bar{x}(t)\| = \|e(t)\| \leq \frac{2\sqrt{n}}{1-\lambda}\epsilon_{avg}(t).$$

■

Lemma 2. *Let $\{a(t)\}_{t\geq 1}$ be a distributed ECFP process as defined in section IV-B (see equations* (14)-(12)*). Then $\|\hat{q}_i(t) - \bar{q}(t)\| = O(\frac{\log t}{t})$, where $\bar{q}(t)$ is the average empirical distribution and $\hat{q}_i(t)$ is player i's estimate of $\bar{q}(t)$.*

Proof: Recall we use the second argument, k, to index the components of the vector $q_i(t) \in \mathbb{R}^m$. Noting that

$$q_i(t+1) = q_i(t) + \frac{1}{t+1}\left(a_i(t+1) - q_i(t)\right),$$

it follows that the maximum incremental change for any single value in the vector $q_i(t)$ is bounded by

$$\begin{aligned} |q_i(t+1,k) - q_i(t,k)| &= |\frac{1}{t+1}\left(a_i(t+1,k) - q_i(t,k)\right)| \\ &\leq \frac{1}{t+1}. \end{aligned}$$

Thus the incremental change in any players empirical distribution is bounded as $|q_i(t+1,k) - q_i(t,k)| \leq \epsilon(t)$, where $\epsilon(t) = \frac{1}{t+1}$. Note that the distributed ECFP process (14) is updated column-wise (each column corresponds to an action k) using an update rule equivalent to (17) of Lemma 1. Also note that, column-wise, all necessary conditions of Lemma 1 are satisfied,[8] and specifically, we have $\epsilon(t) = \frac{1}{t+1}$. Thus we apply Lemma 1 column-wise to $\hat{Q}$ and $Q(t)$ of (14), where $x(t)$ of Lemma 1 corresponds to the k'th column of $Q(t)$, and $\hat{X}(t)$ of Lemma 1 corresponds to the k'th column of $\hat{Q}(t)$, and obtain

$$\|\begin{pmatrix} \hat{q}_1(t,k) \\ \hat{q}_2(t,k) \\ \vdots \\ \hat{q}_n(t,k) \end{pmatrix} - \bar{q}(t,k)\mathbf{1}\| \leq \frac{2\sqrt{n}}{1-\lambda}\epsilon_{avg}(t) = O\left(\frac{\log t}{t}\right),$$

where $\epsilon_{avg}(t) = \frac{1}{t}\sum_{s=1}^{t}\frac{1}{t+1} = O\left(\frac{\log t}{t}\right)$. Thus, $|\hat{q}_i(t,k) - \bar{q}(t,k)| = O\left(\frac{\log t}{t}\right),\ k = 1,\dots,m,\ \forall i$, and hence

$$\|\hat{q}_i(t) - \bar{q}(t)\| = O(\frac{\log t}{t}),\ \forall i.$$

∎

B. Intermediate Results

Lemma 3. *Suppose the sum $-\infty < \sum_{t=1}^{\infty}\frac{a_t}{t} = S < \infty$ converges, then $\lim_{T\to\infty}\frac{a_1+a_2+\ldots+a_T}{T} = 0$.*

[8]The assumption of zero initial error (**A.10**) is satisfied since the initialization of $\hat{q}_i(1)$ in (11) is equivalent to letting $\bar{q}_i(0) = 0$, $\hat{q}_i(0) = 0$ for all i in (13).

Proof: By Kronecker's Lemma [44], $-\infty < \sum_{k=1}^{\infty} \frac{a_k}{k} = S < \infty \Rightarrow \lim_{T\to\infty} \frac{1}{T} \sum_{k=1}^{T} k \frac{a_k}{k} = 0$, which implies that

$$\lim_{T\to\infty} \frac{a_1 + \ldots + a_t}{T} = 0.$$

■

Lemma 4. *For $i \in N$, let $\{q_{-i}(t)\}_{t=1}^{\infty} \in \Delta_{-i}$ and $\{r_{-i}(t)\}_{t=1}^{\infty} \in \Delta_{-i}$ be sequences such that $\|q_{-i}(t) - r_{-i}(t)\| = O\left(\frac{\log t}{t^r}\right),\ r > 0$. Let $U(p) : \Delta^n \to \mathbb{R}$ be the (multilinear) mixed utility function defined in* (1). *Then*

$$|\max_{p_i\in\Delta_i} U(p_i, q_{-i}(t)) - \max_{p_i\in\Delta_i} U(p_i, r_{-i}(t))| = O\left(\frac{\log t}{t^r}\right).$$

Proof: Let $\zeta'_{-i} \in \Delta_{-i}$ and $\zeta''_{-i} \in \Delta_{-i}$. Let $p^* = \arg\max_{p_i\in\Delta_i} U(p_i, \zeta'_{-i})$ and $p^{**} = \arg\max_{p_i\in\Delta_i} U(p_i, \zeta''_{-i})$.[9] $U(\cdot)$ is multilinear and is therefore Lipschitz continuous over the domain Δ^n. Let K be the Lipschitz constant for $U(\cdot)$ such that $|U(x) - U(y)| \leq K\|x - y\|$ for $x, y \in \Delta^n$. By Lipschitz continuity, it holds that

$$\begin{aligned} U(p^*, \zeta'_{-i}) &\leq U(p^*, \zeta''_{-i}) + K\|\zeta'_{-i} - \zeta''_{-i}\| \\ &\leq U(p^{**}, \zeta''_{-i}) + K\|\zeta'_{-i} - \zeta''_{-i}\|, \end{aligned} \tag{21}$$

and thus $U(p^*, \zeta'_{-i}) - U(p^{**}, \zeta''_{-i}) \leq K\|\zeta'_{-i} - \zeta''_{-i}\|$. By a symmetric argument to (21), we also establish $U(p^{**}, \zeta''_{-i}) - U(p^*, \zeta'_{-i}) \leq K\|\zeta'_{-i} - \zeta''_{-i}\|$, thus

$$|U(p^*, \zeta'_{-i}) - U(p^{**}, \zeta''_{-i})| \leq K\|\zeta'_{-i} - \zeta''_{-i}\|.$$

From the above it follows that,

$$\begin{aligned} &|\max_{p_i\in\Delta_i} U(p_i, q_{-i}(t)) - \max_{p_i\in\Delta_i} U(p_i, r_{-i}(t))| \\ &\qquad \leq K\|q_{-i}(t) - r_{-i}(t)\|, \end{aligned}$$

implying the result

$$|\max_{p_i\in\Delta_i} U(p_i, q_{-i}(t)) - \max_{p_i\in\Delta_i} U(p_i, r_{-i}(t))| = O\left(\frac{\log t}{t^r}\right).$$

■

[9] Note that such p^* and p^{**} exist, as $U(\cdot)$ is continuous and the maximization set is compact.

Lemma 5. *Suppose* $|a_t - b_t| = O(\frac{\log t}{t^r}), r > 0$, $b_t \geq 0$ *and* $\sum_{t=1}^{T} \frac{a_t}{t} < \overline{B}$ *is bounded above by* $\overline{B} \in \mathbb{R}$ *for all* $T > 0$ *. Then* $\sum_{t=1}^{T} \frac{b_t}{t}$ *converges as* $T \to \infty$*.*

Proof: Let

$$\delta_t = \begin{cases} b_t - a_t & \text{if } b_t > a_t \\ 0 & \text{otherwise.} \end{cases}$$

It follows that $\delta_t \geq 0$ and $b_t \leq a_t + \delta_t$. By hypothesis, $|a_t - b_t| = O(\frac{\log t}{t^r})$, which implies that $\delta_t = O(\frac{\log t}{t^r})$. It follows that,

$$\begin{aligned} \sum_{t=1}^{T} \frac{b_t}{t} &\leq \sum_{t=1}^{T} \frac{a_t + \delta_t}{t} \\ &= \sum_{t=1}^{T} \frac{a_t}{t} + \sum_{t=1}^{T} \frac{\delta_t}{t}. \end{aligned}$$

Since $\sum\limits_{t=1}^{\infty} \frac{a_t}{t} < \overline{B}$ is bounded above, $\sum\limits_{t=1}^{\infty} \frac{\delta_t}{t} < \infty$ converges, and $b_t \geq 0$, it follows that $\sum\limits_{t=1}^{T} \frac{b_t}{t} < \infty$ converges as $T \to \infty$. ∎

Lemma 6. *Let* $a_t = \sum\limits_{i=1}^{n} [v_i^m(\bar{q}(t)) - U(\bar{q}^n(t))]$, *then* $\lim\limits_{T\to\infty} \frac{a_1 + \ldots + a_T}{T} = 0$ *implies that, for every* $\varepsilon > 0$*,*

$$\lim_{T\to\infty} \frac{\#\{1 \leq t \leq T : \bar{q}^n(t) \notin C_\varepsilon\}}{T} = 0.$$

Proof: Let $\varepsilon > 0$ be given. By definition,

$$\bar{q}^n(t) \in C_\varepsilon \Leftrightarrow v_i^m(\bar{q}(t)) - U(\bar{q}^n(t)) < \varepsilon \ \forall i. \tag{22}$$

The utility function $U(\cdot)$ is assumed to be permutation invariant for all players, so an equivalent statement to (22) is,

$$\bar{q}^n(t) \in C_\varepsilon \Leftrightarrow \sum_{i=1}^{n} [v_i^m(\bar{q}(t)) - U(\bar{q}^n(t))] < n\varepsilon.$$

Let

$$b_t = \begin{cases} 1, & \text{if } a_t \geq n\varepsilon \\ 0, & \text{otherwise.} \end{cases}$$

Note that $b_t = 0 \Leftrightarrow \bar{q}^n(t) \in C_\varepsilon$ and $b_t = 1 \Leftrightarrow \bar{q}^n(t) \notin C_\varepsilon$, thus

$$\frac{\#\{1 \leq t \leq T : \bar{q}^n(t) \notin C_\varepsilon\}}{T} = \frac{b_1 + \ldots + b_T}{T}.$$

Note also that $a_t \geq 0$. Clearly,

$$\frac{b_1 + \ldots + b_T}{T} \leq \frac{1}{n\varepsilon} \frac{a_1 + \ldots + a_T}{T},$$

implying $\lim_{T \to \infty} \frac{b_1 + \ldots + b_T}{T} = 0$, from which the desired result,

$$\lim_{t \to \infty} \frac{\#\{1 \leq t \leq T : \bar{q}^n(t) \notin C_\varepsilon\}}{T} = 0$$

follows. ■

Lemma 7. *Let $\bar{p}^n = (\bar{p}_1, \bar{p}_2, \ldots, \bar{p}_n)$ be a consensus distribution such that $\bar{p}_1 = \bar{p}_2 \ldots = \bar{p}_n$ and let $\delta > 0$ be given. Then there exists an $\varepsilon > 0$ such that $\bar{p}^n \notin B_\delta(C)$ implies $\bar{p}^n \notin C_\varepsilon$.*

Proof: Let $B_\delta^c(C)$ denote the complement of the set $B_\delta(C)$. Suppose for the sake of contradiction that there does not exist an $\varepsilon > 0$ such that $\bar{p}^n \notin B_\delta(C) \Rightarrow \bar{p}^n \notin C_\varepsilon$. Then there exists some $\bar{p}^n \in B_\delta^c(C)$ such that $\bar{p}^n \in C_\varepsilon \ \forall \varepsilon > 0$. This implies that $\bar{p}^n \in C$, or equivalently, $d(\bar{p}^n, C) = 0$. But the hypothesis $\bar{p}^n \in B_\delta^c(C)$ implies that $d(\bar{p}^n, C) \geq \delta$, a contradiction. ■

Lemma 8. $\lim_{T \to \infty} \frac{\#\{1 \leq t \leq T : \bar{q}^n(t) \notin C_\varepsilon\}}{T} = 0$ *for all* $\varepsilon > 0$ *implies that* $\lim_{T \to \infty} \frac{\#\{1 \leq t \leq T : \bar{q}^n(t) \notin B_\delta(C)\}}{T} = 0$ *for all* $\delta > 0$.

Proof: Suppose

$$\lim_{T \to \infty} \frac{\#\{1 \leq t \leq T : \bar{q}^n(t) \notin C_\varepsilon\}}{T} = 0$$

for all $\varepsilon > 0$, but there exists some $\delta > 0$ such that $\limsup_{T \to \infty} \frac{\#\{1 \leq t \leq T : \bar{q}^n(t) \notin B_\delta(C)\}}{T} = \alpha > 0$. By Lemma 7, there exists an $\varepsilon' > 0$ such that $\bar{q}^n(t) \notin B_\delta(C) \Rightarrow \bar{q}^n(t) \notin C_{\varepsilon'}$, which implies that

$$\begin{aligned} &\#\{1 \leq t \leq T : \bar{q}^n(t) \notin C_{\varepsilon'}\} \\ &\geq \#\{1 \leq t \leq T : \bar{q}^n(t) \notin B_\delta(C)\}. \end{aligned}$$

Implying

$$\limsup_{T \to \infty} \frac{\#\{1 \leq t \leq T : \bar{q}^n(t) \notin C_{\varepsilon'}\}}{T} \geq \alpha$$

for some $\varepsilon' > 0$, a contradiction. ■

Lemma 9. $\lim_{T \to \infty} \frac{\#\{1 \leq t \leq T : \bar{q}^n(t) \notin B_\delta(C)\}}{T} = 0$ *for all* $\delta > 0$ *implies* $\lim_{t \to \infty} d(\bar{q}^n(t), C) = 0$.

Proof: Our proof follows the methodology of [7], but is adapted to show convergence to the set of consensus equilibria. Let $\{a(t)\}_{t=1}^\infty$ be an empirical centroid fictitious play process, and let $\{\bar{q}^n(t)\}_{t=1}^\infty$ be

the associated average belief process. Let $\delta > 0$ be given, let $M = \max_{p',p'' \in \Delta^n} \|p' - p''\|$, and let $\eta < \frac{\delta}{2\delta + M}$. The hypothesis $\lim_{T\to\infty} \frac{\#\{1 \le t \le T : \bar{q}^n(t) \notin B_\delta(C)\}}{T} = 0$ implies that there exists an integer T_0 such that for every $T \ge T_0$

$$\#\{1 \le t \le T : \bar{q}^n(t) \notin B_\delta(C)\} < \eta T. \tag{23}$$

We claim that for every $T \ge T_0$, $\bar{q}^n(T) \in B_{2\delta}(C)$.

Suppose $T \ge T_0$ and $\bar{q}^n(T) \notin B_{2\delta}(C)$. Then for $T < t < T + \frac{\delta}{\delta+M}T$, $\bar{q}^n(t) \notin B_\delta(C)$. In order to verify this note that,

$$\|\bar{q}^n(T+1) - \bar{q}^n(T)\| \le \frac{M}{T+1}$$

and for $T \le t \le T + \frac{\delta}{\delta+M}T$

$$\begin{aligned}
\|\bar{q}^n(t) - \bar{q}^n(T)\| &\le M \sum_{s=T}^{\lfloor T + \frac{\delta}{\delta+M}T \rfloor} \frac{1}{s} \\
&\le M \frac{\delta}{\delta + M} T \frac{1}{T} \\
&\le \delta \frac{M}{\delta + M} \\
&\le \delta.
\end{aligned}$$

Since $\bar{q}^n(t) \notin B_\delta(C)$ for $T \le t \le T + \frac{\delta}{\delta+M}T$, we have

$$\begin{aligned}
&\#\{1 \le t \le T + \frac{\delta}{\delta + M}T : \bar{q}^n(t) \notin B_\delta(C)\} \ge \frac{\delta}{\delta + M}T \\
&= \frac{\delta}{2\delta + M}\left(T + \frac{\delta}{\delta + M}T\right) > \eta \left(T + \frac{\delta}{\delta + M}T\right),
\end{aligned}$$

contradicting (23). Therefore, for any $\delta > 0$, there exists a T_0 such that, for all $T \ge T_0$, $d(\bar{q}^n(T), C) < \delta$, i.e. $\lim_{t\to\infty} d(\bar{q}^n(t), C) = 0$. ■

Lemma 10. *Let $\{a_t\}_{t=1}^T,\ a_t \ge 0$ be a monotone non-increasing sequence and let $\{b_t\}_{t=1}^T,\ b_t \ge 0$ be a monotone non-decreasing sequence. Let $b_{avg} = \frac{1}{T}\sum_{t=1}^T b_t$ be the mean of $\{b_t\}_{t=1}^T$. Then $\sum_{t=1}^T a_t b_t \le b_{avg} \sum_{t=1}^T a_t$.*

Proof: We represent the sequences using vectors $a, b \in \mathbb{R}^T$ and prove the result in this framework. Let $b(1) \le b(2) \le \ldots \le b(T)$ and $a(1) > a(2) > \ldots > a(T)$. We will prove the result for a, strictly increasing first and then show that it generalizes to the non-decreasing case. Assume without loss of

generality that $\mathbf{1}^T a = \mathbf{1}^T b = 1$. Consider the optimization problem

$$\max_{\substack{\beta \in \mathbb{R}^T \\ \beta \geq 0 \\ \mathbf{1}^T \beta = 1 \\ \beta_1 \leq \ldots \leq \beta_T}} \quad a^T \beta \tag{24}$$

The solution to (24) is $\beta^* = \frac{1}{T}\mathbf{1}$. To see this, assume for the sake of contradiction that $\beta^* = \frac{1}{T}\mathbf{1}$ is not a solution. Let β be a solution such that $\beta(i+1) < \beta(i)$. Create a new vector $\hat{\beta}$ which is a duplicate of β, with the exception that,

$$\hat{\beta}(i+1) = \hat{\beta}(i) = \frac{\beta(i+1) - \beta(i)}{2}.$$

Then

$$\begin{aligned} a^T \hat{\beta} - a^T \beta &= a_i \left(\hat{\beta}(i) - \beta(i) \right) + a_{i+1} \left(\hat{\beta}(i+1) - \beta(i+1) \right) \\ &= a(i)(\delta) + a(i+1)(-\delta) \\ &= \delta \left(a(i) - a(i+1) \right) \\ &> 0, \end{aligned}$$

a contradiction. We now consider relaxing the assumption on a; we let a be non-decreasing. Let $\{a_t\}_{t=1}^{\infty}, a_t \in \mathbb{R}^T$ be an sequence of approximations of a such that a_t is strictly increasing $(a_t(1) < a_t(2) < \ldots a_t(T))$ and $\lim_{t \to \infty} a - a_t = 0$. Then we have $a_t^T b \leq a_t^T \frac{1}{T}\mathbf{1}$. Taking the limit as $t \to \infty$,

$$a^T b \leq a^T \frac{1}{T}\mathbf{1} = \frac{1}{T}\mathbf{1}^T b \mathbf{1}^T a = b_{avg}\mathbf{1}^T a.$$

■